(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété Intellectuelle

Bureau international



(43) Date de la publication internationale 27 octobre 2005 (27.10.2005)

- (51) Classification internationale des brevets⁷: G06K 9/00
- (21) Numéro de la demande internationale :

PCT/FR2005/000673

- (22) Date de dépôt international : 18 mars 2005 (18.03.2005)
- (25) Langue de dépôt :

français

(26) Langue de publication :

français

FR

- (30) Données relatives à la priorité : 0403556 5 avril 2004 (05.04.2004)
- (71) Déposant et
- (72) Inventeur: LEBRAT, François [FR/FR]; 98 avenue de Versailles, F-75016 Paris (FR).

(74) Mandataires: HASSINE, Albert etc.; Cabinet Plasseraud, 65/67, rue de la Victoire, F-75440 Paris Cedex 09 (FR).

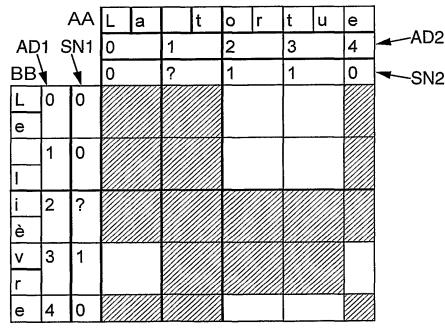
WO 2005/101292 A3

- (81) États désignés (sauf indication contraire, pour tout titre de protection nationale disponible): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) États désignés (sauf indication contraire, pour tout titre de protection régionale disponible): ARIPO (BW, GH,

[Suite sur la page suivante]

(54) Title: METHOD FOR SEARCHING CONTENT PARTICULARLY FOR EXTRACTS COMMON TO TWO COMPUTER FILES

(54) Titre : PROCÉDÉ DE RECHERCHE DE CONTENU, NOTAMMENT D'EXTRAITS COMMUNS ENTRE DEUX FICHIERS INFORMATIQUES



AA...THE TORTOISE BB...THE HARE

(57) Abstract: The present invention relates searching to content, particularly for at least one extract common to a first data file and a second data file. method comprises a preliminary step of preparing at least the first file by (a) dividing the first file into a series of data packets having a predetermined size, and identifying packet addresses in said file, (b) combining each packet address with a digital signature that defines one of three fuzzy logic states, namely true, false and indeterminate, and is the result of a combinatorial computation on data from said file; whereafter said method comprises performing an actual search for a common extract by (c) comparing the fuzzy logic states combined with each packet address of the first file with fuzzy logic states determined on the basis of data from the second file, and (d) removing from said common extract search the respective address pairs from the first and second files that have the respective logic states true and false or false and true, and

retaining the other address pairs that identify data packets that may comprise said common extract.

WO 2005/101292 A3



GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Déclaration en vertu de la règle 4.17 :

 relative à la qualité d'inventeur (règle 4.17.iv)) pour US seulement

Publiée:

avec rapport de recherche internationale

 avant l'expiration du délai prévu pour la modification des revendications, sera republiée si des modifications sont reçues

(88) Date de publication du rapport de recherche internationale: 26 janvier 2006

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

(57) Abrégé: L'invention concerne la recherche de contenu, notamment une recherche d'au moins un extrait commun entre un premier fichier et un second fichier de données. Le procédé de l'invention comporte une préparation préalable du premier fichier au moins, comprenant les étapes suivantes: a) segmenter le premier fichier en une succession de paquets de données, de taille choisie, et identifier des adresses de paquets dans ledit fichier, b) associer à l'adresse de chaque paquet une signature numérique définissant un état en logique floue parmi au moins trois états: "vrai", "faux" et "indéterminé", ladite signature résultant d'un calcul combinatoire sur des données issues dudit fichier, et le procédé se poursuit par une recherche d'extrait commun, proprement dite, comprenant les étapes suivantes, c) comparer les états de logique floue associés à chaque adresse de paquet du premier fichier, avec des états de logique floue déterminés à partir de données issues du second fichier, d) éliminer de ladite recherche d'extrait commun des couples d'adresses respectives des premier et second fichiers dont les états logiques respectifs sont "vrai" et "faux" ou "faux" et "vrai", et conserver les autres couples d'adresses identifiant des paquets de données susceptibles de comporter ledit extrait commun.

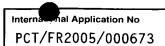
INTERNATIONAL SEARCH REPORT

Internation Application No	
PCT/FR2005/000673	

A CLASSI IPC 7	FICATION OF SUBJECT MATTER G06K9/00		
According to	o International Patent Classification (IPC) or to both national classification	ation and IPC	
B. FIELDS	SEARCHED		
Minimum do IPC 7	ocumentation searched (classification system followed by classification ${\tt G06K-G06F-H04L}$	on symbols)	
	tion searched other than minimum documentation to the extent that s		
Electronic da	ata base consulted during the international search (name of data ba	se and, where practical, search terms used	'
EPO-In	ternal, INSPEC, IBM-TDB, PAJ, COMPEN	IDEX	
C. DOCUME	ENTS CONSIDERED TO BE RELEVANT		
Category °	Citation of document, with indication, where appropriate, of the rele	evant passages	Relevant to claim No.
Т	SNASEL V: "Fuzzy signatures for multimedia databases" ADVANCES IN INFORMATION SYSTEMS. FIRST INTERNATIONAL CONFERENCE, ADVIS 2000. PROCEEDINGS (LECTURE NOTES IN COMPUTER SCIENCE VOL.1909) SPRINGER VERLAG BERLIN, GERMANY, 2000, pages 257-264, XP008038840 ISBN: 3-540-41184-4 p. 259-260, paragraphe "4. fuzzy Signatures"		1-30
X Furth	ner documents are listed in the continuation of box C.	Patent family members are listed in	annex.
 Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone document is combined with one or more other such document is combined with one or more other such document is combination being obvious to a person skilled in the art. "&" document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention cannot be considered novel or cannot be considered to involve an inventive step when the document is combined with one or more other such document is combined with one or more other such document is combination being obvious to a person skilled in the art. "&" document member of the same patent family 			the application but ory underlying the aimed invention be considered to cument is taken alone aimed invention rentive step when the re other such docusto a person skilled
Date of the a	actual completion of the international search	Date of mailing of the international sear	ch report
7	November 2005	14/11/2005	
Name and m	nailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl,	Authorized officer Granger, B	

2

INTERNATIONAL SEARCH REPORT



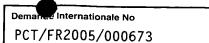
		PC1/FR2005/0006/3
C.(Continu	ation) DOCUMENTS CONSIDERED TO BE RELEVANT	
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	S. HABER ET AL: "If Piracy is the Problem, Is DRM the Answer?" DIGITAL RIGHTS MANAGEMENT. TECHNOLOGICAL, ECONOMIC, LEGAL AND POLITICAL ASPECTS, 31 August 2003 (2003-08-31), pages 224-233, XP002307417 HEIDELBERG p. 229, paragraphe "III.4 Fuzzy Hashing"	1-30
Г	FALOUTSOS C: "ACCESS METHODS FOR TEXT" ACM COMPUTING SURVEYS, NEW YORK, NY, US, vol. 17, no. 1, March 1985 (1985-03), pages 49-74, XP000747611 ISSN: 0360-0300 p.54-56, paragraphe "2.3 Multiattribute Retrieval Methods"	1-30
		•
	·	

RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No
PCT/FR2005/000673

			PC1/FR2005/0006/3
CIB 7	EMENT DE L'OBJET DE LA DEMANDE G06K9/00		
	assification internationale des brevets (CIB) ou à la fois selon la class	ification nationale et la CIE	<u>. </u>
	NES SUR LESQUELS LA RECHERCHE A PORTE		
CIR /	ition minimale consultée (système de classification suivi des symbole G06K G06F H04L		
	tion consultée autre que la documentation minimale dans la mesure		
	nnées électronique consultée au cours de la recherche internationale ternal, INSPEC, IBM-TDB, PAJ, COMPEN		ees, et si realisable, termes de recherche utilisés)
C. DOCUM	ENTS CONSIDERES COMME PERTINENTS		
Catégorie °	Identification des documents cités, avec, le cas échéant, l'indication	n des passages pertinents	no. des revendications visées
T	SNASEL V: "Fuzzy signatures for multimedia databases" ADVANCES IN INFORMATION SYSTEMS. INTERNATIONAL CONFERENCE, ADVIS 2 PROCEEDINGS (LECTURE NOTES IN COM SCIENCE VOL.1909) SPRINGER VERLAG GERMANY, 2000, pages 257-264, XP0 ISBN: 3-540-41184-4 p. 259-260, paragraphe "4. fuzzy Signatures"	2000. IPUTER BERLIN	1-30
X Voir la	a suite du cadre C pour la fin de la liste des documents	Les documents de	familles de brevets sont indiqués en annexe
"A" documen considé "E" documen ou aprè: "L" documen priorité dautre ci "O" documen une exp "P" documen	nt définissant l'état général de la technique, non ré comme particulièrement pertinent it antérieur, mais publié à la date de dépôt international s cette date "; t pouvant jeter un doute sur une revendication de pu cité pour déterminer la date de publication d'une aitoin ou pour une raison spéciale (telle qu'indiquée) it se référant à une divulgation orale, à un usage, à osition ou tous autres moyens it publié ayant la date de dépôt international, mais	technique pertinent, m ou la théorie constitua X" document particulièrem être considérée comm inventive par rapport a Y" document particulièrem ne peut être considéré lorsque le document e documents de même r pour une personne du	lié après la date de dépôt international ou la partenenant pas à l'état de la ais cité pour comprendre le principe nt la base de l'invention ent pertinent; l'inven tion revendiquée ne peut e nouvelle ou comme impliquant une activité u document considéré isolément ent pertinent; l'inven tion revendiquée e comme impliquant une activité inventive st associé à un ou plusieurs autres nature, cette combinaison étant évidente métier de la même famille de brevets
Date à laquell	e la recherche internationale a été effectivement achevée		ésent rapport de recherche internationale
	novembre 2005	14/11/200!	5
Nom et adress	se postale de l'administration chargée de la recherche internationale Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Fonctionnaire autorisé Granger, E	3

RAPPORT DE RECHERCHE INTERNATIONALE



Catégorie °	S. HABER ET AL: "If Piracy is the Problem, Is DRM the Answer?" DIGITAL RIGHTS MANAGEMENT. TECHNOLOGICAL,	tinents	no. des revendications visées 1-30
	S. HABER ET AL: "If Piracy is the Problem, Is DRM the Answer?" DIGITAL RIGHTS MANAGEMENT. TECHNOLOGICAL	tinents	
T	Problem, Is DRM the Answer?" DIGITAL RIGHTS MANAGEMENT, TECHNOLOGICAL		1-30
	ECONOMIC, LEGAL AND POLITICAL ASPECTS, 31 août 2003 (2003-08-31), pages 224-233, XP002307417 HEIDELBERG p. 229, paragraphe "III.4 Fuzzy Hashing"		1 30
Т	FALOUTSOS C: "ACCESS METHODS FOR TEXT" ACM COMPUTING SURVEYS, NEW YORK, NY, US, vol. 17, no. 1, mars 1985 (1985-03), pages 49-74, XP000747611 ISSN: 0360-0300 p.54-56, paragraphe "2.3 Multiattribute Retrieval Methods"		1-30
			,
٠			